10К

УДК 517.958:531.33

Ю.В. Крак[1], І.О. Стеля[2]
Київський національний університет імені Тараса Шевченка, м. Київ, Україна
[1]krak@unicyb.kiev.ua
[2]igor.stelia@gmail.com

Синтез звуків голосу людини на основі фізичних моделей голосових зв'язок та мовного тракту

Стаття присвячена проблемі синтезу мовного сигналу на основі фізичних моделей голосових зв'язок та мовного тракту. Запропоновано математичне та програмне забезпечення для реалізації цих моделей. Проведена серія чисельних експериментів.

Вступ

Метою роботи є створення алгоритмів та програмного забезпечення для відтворення звуків голосу людини з використанням фізичних моделей голосових зв'язок та мовного тракту. В більшості публікацій на цю тему вивчаються або моделі голосових зв'язок, або моделі мовного тракту. В даній роботі результат моделювання голосового джерела – потік повітря на виході з голосової щілини – використовується як вихідні дані в моделі мовного тракту. Здійснений порівняльний аналіз результатів моделювання з аналітичними моделями. Згідно з розробленими алгоритмами створено програмне забезпечення та проведено обчислювальний експеримент із синтезу звуків.

Дослідження моделі голосових зв'язок

Для дослідження обрана двомасова модель [1] голосових зв'язок, відповідно до якої кожна з голосових зв'язок описується двома масами, що здійснюють автоколивання під впливом потоку повітря в голосовій щілині. В основу алгоритму для реалізації моделі покладений метод Кутта – Мерсона та метод нелінійної релаксації, які об'єднані в єдину ітераційну процедуру [2] для розв'язування системи, що описує зміщення мас:

$$\begin{aligned} m_1\frac{d^2x_1(t)}{dt^2}+r_1\frac{dx_1(t)}{dt}+k_1(x_1(t)-x_{01})+k_c(x_1(t)-x_2(t))=l_g d_1 p_{m1}(t),\\ m_2\frac{d^2x_2(t)}{dt^2}+r_2\frac{dx_2(t)}{dt}+k_2(x_2(t)-x_{02})-k_c(x_1(t)-x_2(t))=l_g d_2 p_{m2}(t), \end{aligned} \quad (1)$$

де m_1 і m_2 – маси, $x_1(t)$, $x_2(t)$ – зміщення, відповідно, мас m_1 і m_2, x_{01},x_{02} – положення мас у стані спокою, t – час, r_1 і r_2 – коефіцієнти демпфування, k_1, k_2 – пружність пружин для мас m_1 і m_2, відповідно, k_c – пружність пружини, що з'єднує

маси, d_1, d_2 – товщина мас, l_g – діюча довжина голосових зв'язок, $l_g d_1$, $l_g d_2$ – поверхні мас, на які діють тиски, відповідно, $p_{m1}(t)$ і $p_{m2}(t)$, та нелінійної системи для змін тиску:

$$p_s - p_{11}(t) = 0.69\rho \frac{u_g^2(t)}{A_{g1}^2(t)} + \int_0^{l_c} \frac{\rho}{A_c(x)} dx \cdot \frac{du_g}{dt},$$

$$p_{11}(t) - p_{12}(t) = 12\nu d_1 \frac{l_g^2 u_g(t)}{A_{g1}^3(t)} + \frac{\rho d_1}{A_{g1}} \cdot \frac{du_g}{dt},$$

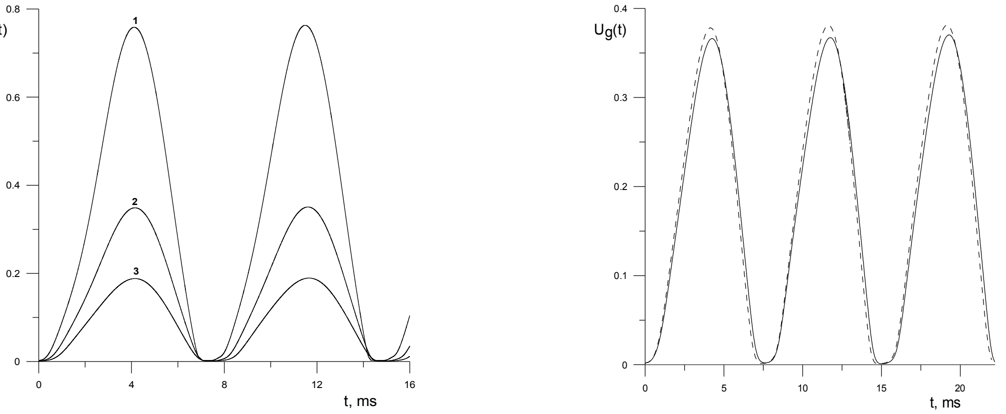
$$p_{12}(t) - p_{21}(t) = \frac{1}{2}\rho u_g^2(t)\left(\frac{1}{A_{g2}^2(t)} - \frac{1}{A_{g1}^2(t)}\right), \qquad (2)$$

$$p_{21}(t) - p_{22}(t) = 12\nu d_2 \frac{l_g^2 u_g(t)}{A_{g2}^3(t)} + \frac{\rho d_2}{A_{g2}} \cdot \frac{du_g}{dt},$$

$$p_{22}(t) - p = -\frac{1}{2}\rho \frac{u_g^2(t)}{A_{g2}^2(t)}\left[2\frac{A_{g2}(t)}{A_1}(1 - \frac{A_{g2}(t)}{A_1})\right],$$

де ρ – густина повітря, ν – зсувова в'язкість повітря, l_c – довжина звуження, $A_c(x)$ – площа поперечного перетину звуження, A_1 – площа голосового тракту на вході, A_{gi} – площа голосової щілини під i-ю масою. Шуканим розв'язком задачі (1) – (2) є функція u_g, яка визначає потік повітря на виході з голосової щілини. Цей потік використовується як вихідні дані для моделі мовного тракту.

Проведена серія чисельних експериментів, за результатами яких здійснений аналіз впливу параметрів та інерційних членів моделі на форму вихідного сигналу, від якого великою міру залежить частота та тембр голосу. Зокрема, залежність потоку повітря u_g від тиску p_s на вході у голосову щілину зображена на рис. 1а. Вплив інерційних членів моделі зображений на рис. 1б.



а) б)

Рисунок 1 – а) Результати розрахунків потоку повітря $u_g(t)$ для трьох значень тиску на вході в голосову щілину $p_{s_1} = 0.02$, $p_{s_2} = 0.0012$, $p_{s_3} = 0.008$; б) потік повітря u_g на виході з голосових зв'язок, обчислений з інерційними членами (суцільна лінія) та без них (пунктирна лінія)

Здійснене порівняння потоку повітря, обчисленого за двомасовою моделлю, та потоку повітря, отриманого за аналітичною R-моделлю [3]:

$$u_g(t) = \begin{cases} \alpha\left(3\left(\frac{t}{t_p}\right)^2 - 2\left(\frac{t}{t_p}\right)^3\right), 0 \le t \le t_p, \\ \alpha\left(1 - \left(\frac{t - t_p}{t_p}\right)^2\right), t_p \le t \le t_p + t_n \le t_0, \end{cases}$$

де α – амплітуда, t_p – час відкриття, t_n – час закриття, t_0 – період. Результати порівняння моделей наведені на рис. 2.

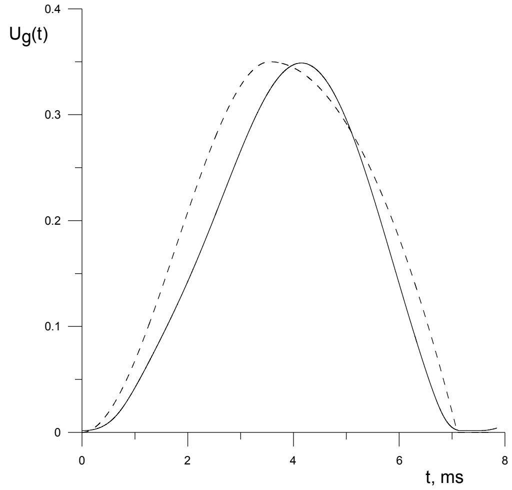


Рисунок 2 – Потік повітря u_g на виході з голосових зв'язок, обчислений за двомасовою моделлю (суцільна лінія) та R-моделлю (пунктирна лінія)

Дослідження похідної потоку повітря

Для вивчення голосового джерела, окрім функції потоку повітря, використовується її похідна. Звичайний набір параметрів для кількісного опису коливань виводиться із залежності похідної потоку від часу і включає [4]: амплітуду від'ємного піка, тривалість відкриття, тривалість закритої фази, проміжок часу між додатним і від'ємним піками, час повернення в початкове положення і вторинні параметри, що походять з них: швидкість відкриття, тривалість фази відкриття голосової щілини відносно періоду основного тону, тривалість фази відкритої голосової щілини відносно тривалості основного тону тощо.

Проведений порівняльний аналіз отриманої чисельної похідної потоку повітря за двомасовою моделлю та похідної потоку повітря, отриманою за аналітичною LF-моделлю [5]:

$$\dot{u}_g(t) = \begin{cases} E_0 e^{\alpha t} \sin(\omega_g t), \quad 0 \le t < t_e, \\ -\frac{E_e}{\varepsilon t_a}\left[e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c - t_e)}\right], \quad t_e \le t < t_c, \\ 0, \quad t_c \le t < t_0, \end{cases}$$

де t_0 – період, t_e – відмітка мінімального значення похідної, t_a – визначається як точка, в якій тангенс перетинає вісь координат, t_c – момент, в який похідна під час фази закриття майже досягає нуля, E_e – абсолютне значення мінімуму похідної, α, ε, ω_g – деякі параметри. Результати порівняння моделей наведені на рис. 3.

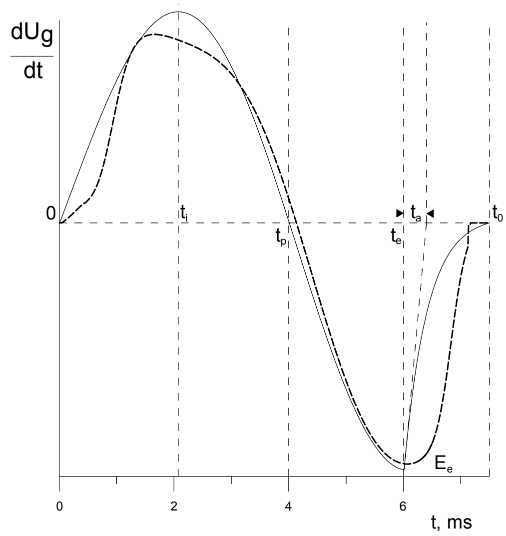


Рисунок 3 – Чисельна похідна від потоку повітря, обчисленого за двомасовою моделлю (пунктирна лінія), та похідна, обчислена за LF-моделлю (суцільна лінія)

Моделювання мовного тракту

Для моделювання розповсюдження акустичних хвиль у мовному тракті як у неоднорідній акустичній трубі, що починається між голосовими зв'язками та закінчується губами, використовується система рівнянь акустики в частинних похідних [6-8], яка записується у вигляді:

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A(x)}\frac{\partial u}{\partial t}, \qquad -\frac{\partial u}{\partial x} = \frac{A(x)}{\rho c^2}\frac{\partial p}{\partial t}, \tag{3}$$

де $0 \le x \le L,\ t > 0$, L – довжина мовного тракту, $p(x,t)$ – тиск у тракті в момент часу t, $u(x,t)$ – об'ємна швидкість потоку, ρ – густина повітря в тракті, c – швидкість звуку і $A(x)$ – функція площі поперечного перетину.

Оскільки тракт має неоднорідний поперечний перетин, він розбивається на циліндричні секції однакової довжини з постійною площею перетину (рис. 4).

Як крайова умова на вході в тракт вибирається потік $u_g(t)$, знайдений з (1) – (2). Виходячи з цього, маємо крайову умову: $u(0,t) = u_g(t)$. На протилежному кінці тракту задаємо умову $p(L,t) = 0$.

Різницева задача для апроксимації системи рівнянь (3) будується на так званій рознесеній сітці. Для розв'язання використовується явний метод «чехарда». Цей алгоритм має високу обчислювальну ефективність і може використовуватися для моделювання в реальному часі [9]. Результати роботи алгоритму наведені на рис. 5.

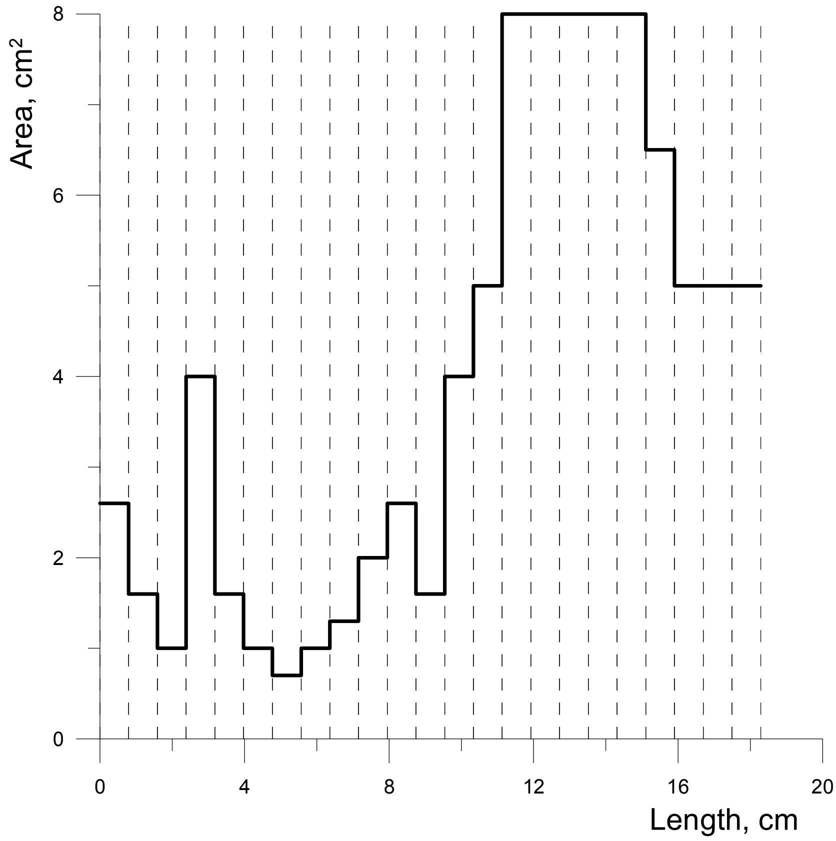


Рисунок 4 – Наближення мовного тракту циліндричними секціями однакової довжини

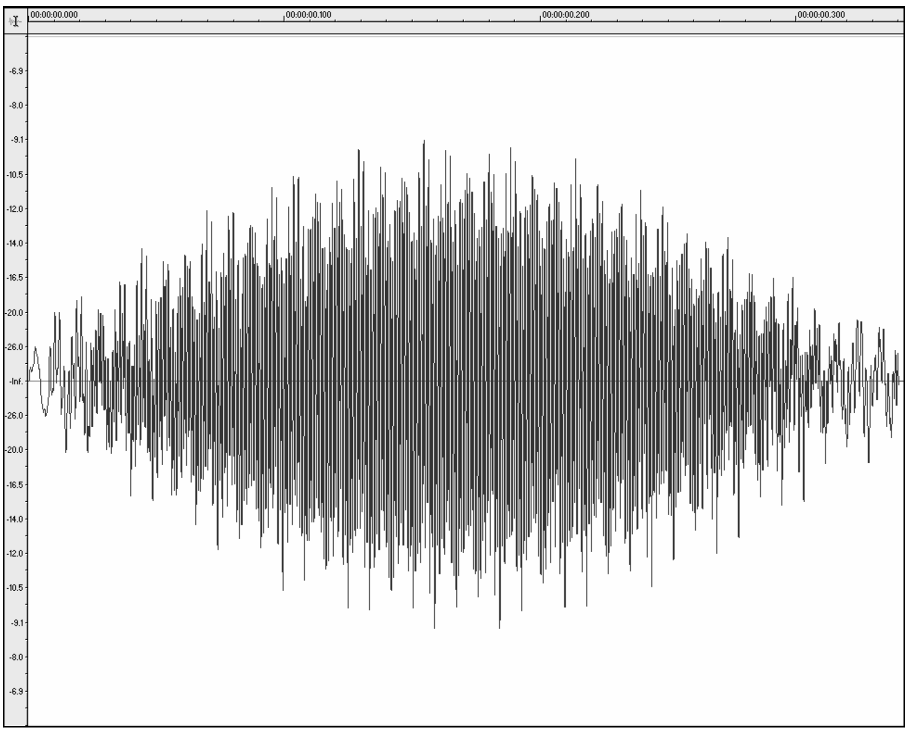


Рисунок 5 – Результати моделювання російського звуку [а]

Висновки

Проведені в роботі дослідження комп'ютерних моделей голосових зв'язок та мовного тракту продемонстрували можливість їх спільного використання в системах артикуляторного синтезу. Порівняння результатів моделювання з аналітичними моделями свідчать про адекватність одержаних результатів, а створене програмне забезпечення дозволяє в зручному режимі задавати різні набори параметрів моделей для генерації звуків голосу людини.

Література

1. Ishizaka K. Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords / K. Ishizaka, J.L. Flanagan // Bell Syst. Tech. J. – 1972. – № 51. – P. 1233-1268.
2. Крак Ю.В. Чисельне моделювання голосових зв'язок за двомасовою моделлю / Ю.В. Крак, І.О. Стеля // Журнал обчислювальної та прикладної математики. – 2007. – № 1(94). – С. 55-60.
3. Rosenberg A. Effect of glottal pulse shape on the quality of natural vowels / A. Rosenberg // Journal of the Acoustical Society of America. – 1971. – № 49. – P. 583-590.
4. Alku P. A comparison of glottal voice source quantification parameters in breathy, normal, and pressed phonation of female and male speakers / P. Alku, E. Vilkman // Folia Phoniatr. Logop. – 1996. – № 48. – P. 240-254.
5. Fant G. The lf-model revisited. Transformations and frequency domain analysis / G. Fant // STL-QPSR, Tech. Rep. – 1995. – № 2 – 3. – P. 119-156.
6. Kinsler, Frey, Coppens, and Saunders Fundamentals of Acoustics // San Diego: Academic Press. – 1982. – ISBN: 9780471029335. – 496 p.
7. Portnoff M.R. A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal-Tract / M.R. Portnoff // S.B. / S.M. thesis, MIT, Cambridge, Mass., 1973.
8. Schroeterand J. Techniques for Estimating Vocal-Tract Shapes from theSpeech Signal / J. Schroeterand, M.M. Sondhi // IEEETrans. Speech Audio Proc. – 1994. – Vol. 2(1). – P. 133-150.
9. K. van den Doel. Real-time numerical solution of Webster's equation on a non-uniform grid / K. van den Doel, U. Ascher // IEEE Trans. Audio, Speech and Language Processing 16 (2008). – P. 1163-1172.

Ю.В. Крак, И.О. Стеля

Синтез звуков голоса человека на основе физических моделей голосовых связок и речевого тракта

Статья посвящена проблеме синтеза речевого сигнала на основе физических моделей голосовых связок и речевого тракта. Предложено математическое и программное обеспечение для реализации этих моделей. Проведена серия числовых экспериментов.

Yu.V. Krak, I.O. Stelia

The Human Voice Synthesis Based on Physical Models of Vocal Folds and a Vocal Tract

This article is devoted to the problem of speech signal synthesizing based on physical models of vocal folds and a vocal tract. The mathematical apparatus and software for implementing of these models were presented. A series of numerical experiments was carried out.

Стаття надійшла до редакції 09.06.2009.